PCT/AU99/00563

09/743533

AU 99/563

Patent Office
Canberra

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES, hereby certify that the annexed is a true copy of the Provisional specification in connection with Application No. PP 4604 for a patent by COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION, GOODMAN FIELDER LIMITED and GROUPE LIMAGRAIN PACIFIC PTY LTD filed on 10 July 1998.

WITNESS my hand this Third
day of August 1999

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND SALES

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

AUSTRALIA

Patents Act 1990

COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION,
GOODMAN FIELDER LIMITED,
GROUPE LIMAGRAIN PACIFIC PTY LTD

PROVISIONAL SPECIFICATION

Invention Title:

Modified proteins

The invention is described in the following statement:

Technical Field

The present invention is directed generally to producing modified proteins and particularly to producing modified proteins for use in food preparations.

Background Art

Wheat storage proteins are classified on the basis of their solubility into two classes. The gliadins are readily soluble in aqueous alcohols and are monomeric proteins with only intramolecular disulphide bonds. The glutenins are present in high molecular weight polymers, stabilised by intermolecular disulfide bonds and are not soluble in aqueous alcohols without reducing agent (Kasarda 1989). These proteins are present in high amount in the endosperm and are considered to act as a store of nitrogen, carbon and sulphur source for seed germination.

Glutenins form a continuous proteinaceous network called gluten. The unique physico-chemical properties of gluten determine the ability of wheat dough to be processed into baked goods (bread, biscuits, cakes), pasta noodles and other food products. It is understood that the glutenins, which form crosslinks with each other through disulfide bonds, are the most important molecules producing the viscoelastic properties of wheat flour dough (MacRitchie 1992). The unique position of wheat in bread making is due to the ability of the dough to retain gas on expansion. The gluten accounts for about 10% of the dough, and consists mainly of proteins (70-80%) together with starch and lipids. Starch could be granular and damaged starch. The lipid reserves of wheat are non-polar, structural and endosperm lipids (Gan et al., 1995). Structural lipids are also called polar lipids. The endosperm lipids are divided into non-starch lipids and starch lysophospholipids. The structure and properties of gluten are determined by molecular interactions and it is important to be understood if the functional properties of gluten are to be manipulated.

A dough results from a large variety of interactions between flour constituents facilitated by water. Starch takes up about 46% of the water. The damaged starch contribute significantly to the water absorption. It has been shown that during hydration, proteins exude visible strands or fibrils. Specific proteins of flour are bound to flour lipids (polar) upon addition of water (Morrison 1989).

Dough development is visualised as a re-orientation of glutenin polymers to form a membrane network with viscoelasticity and gas retaining properties. Covalent (disulfide) and noncovalent (hydrogen, hydrophobic and ionic) bonds are involved in formation of a fully developed dough. Interactions are further modified during fermentation, baking and even after baking. The disulphide bonds of flour proteins play a key role in the interactions in dough. The bonds form relatively strong crosslinks within and between polypeptide chains and also stabilise other less energetic bonds. The disulfide bonds provide the required stability for the protein matrix until the loaf structure is set by the gelatinisation of the starch and the thermal denaturation of the proteins. Hydrogen bonds are considerably weaker than covalent bonds, but contribute significantly to the structure of dough. A unique feature is the ability to interchange with other hydrogen bonds, which facilitate reorientation of protein chains and allow for stress relaxation. Hydrophobic bonds results from nonpolar groups of flour constituents. Because these bonds are reversible, they can readily accommodate viscous flow and thereby facilitate mechanical dough development. Ionic bonds play relatively small part in dough structure formation but some specific components have ionisable part or parts. Therefore ionic bond interactions could be important for the rheological properties (for a review, see Bushuk 1998).

A major limitation to evaluating the contributions of various groups of proteins, and of specific structural features of these molecules, to dough functionality has been the lack of appropriate systems that allow specific proteins to be incorporated and tested within the dough. The situation has recently changed, however, due to two advances. The first is the development of small scale testing equipment (Mixograph, Extensograph) with appropriate procedures for incorporating exogenous proteins, including polymeric glutenins into the dough (Bekes et al., 1994). Advantages of this system are the small amount of proteins required for test and the ability to rapidly test multiple samples produced by, for example, protein engineering. The second recent advance is the development of reliable transformation system for wheat (Weeks et al., 1993, Witrzens et al., 1998), which allows the modification of storage protein composition by the expression of new proteins with, for instance, designed characteristics.

To alter protein-protein, protein-lipid and protein-starch interactions within the gluten matrix, the present inventors have developed a system which enables the incorporation of new surface active molecules or part of molecules into gluten matrix.

Disclosure of Invention

In a first aspect, the present invention consists in a method of producing a modified glutenin or seed-storage protein, the method comprising adding to the glutenin or seed-storage protein a binding domain for a macromolecule such that the glutenin or seed-storage protein has the ability to bind or associate with the macromolecule.

In one preferred form, the glutenin or seed-storage protein further includes one or more cysteine residues incorporated at one or both ends of its amino acid sequence. The addition of the one ore more cysteines allows the modified proteins to be more easily incorporated into gluten in use. The modifications to the glutenin or seed-storage proteins produced according to the present invention allow the incorporation of that protein into the gluten network for food or industrial use.

The present inventors have found that incorporating amino acid sequences (domains) from proteins other than glutenins into glutenin or seed-storage proteins modifies the general properties of gluten when the proteins are used in a range of food applications.

The binding domain can be any domain that will bind ligands that may be useful in food preparation or in food compositions. In a preferred form, the binding domain is a ligand capable of binding lipids or starches. The present inventors have found that the lipid-binding domain of the barley oleosin gene, the lipid-binding regions of the wheat CM16 protein, and the starch-binding domain of the glucoamylase from *Aspergillus niger* are particularly suitable for the present invention. It will be appreciated, however, that other natural or modified domains would also be suitable for the present invention.

One glutenin or seed-storage protein that has been modified by the present inventors is the C hordein gene from barley. It will be appreciated, however, that other glutenin or seed-storage proteins may also be modified according to the present invention. In wheat, such glutenin or seed proteins include low molecular weight glutenins, high molecular weight glutenins, gliadins, puroindolines or grain softness proteins (also known as friabilins),

or Chloroform/Methanol-soluble proteins. Homologues of these proteins exist in other cereals such as diploid, tetraploid and hexaploid wheats, rye, triticale, barley, oats, rice, sorghum, millet and maize and the genes encoding these proteins may also be modified according to the present invention.

In a second aspect, the present invention consists in a modified glutenin or seed-storage protein including one or more a binding domains for a macromolecule added thereto such that the glutenin or seed-storage protein has the ability to bind or associate with the macromolecule.

In a third aspect, the present invention consists in an isolated nucleic acid molecule encoding a modified glutenin or seed-storage protein according to the second aspect of the present invention.

In a fourth aspect, the present invention consists in an isolated nucleic acid molecule according to the third aspect of the present invention incorporated into a cell such that on expression of the nucleic acid molecule, the cell produces the modified glutenin or seed-storage protein.

The cell may be a recombinant bacterial cell for example which is capable of producing the modified glutenin or seed-storage protein. Preferably the bacterial cell is *Escherichia coli*. Alternatively, the cell may be a yeast such as *Picchia* sp. or *Saccharomyces cerevisiae*, an insect cell using an expression system such as the baculovirus expression system, or a mammalian cell. Alternatively, the cell may be a plant cell of a recombinant plant which is capable of producing the modified glutenin or seed-storage protein in the plant's seeds. Preferably the plant cell is a recombinant wheat cell.

In fifth aspect, the present invention consists in the use of a modified glutenin or seed-storage protein according to the second aspect of the present invention in the preparation of a food product.

Examples of food products include leavened or unleavened breads, pasta, noodles, breakfast cereals, snack foods, cakes, pastries or other foods containing flour-based sauces.

The modified glutenin or seed-storage proteins according to the present invention, in use, are capable of modifying the structure of doughs and other materials containing gluten in ways which add value and utility to the resultant product. The modified glutenin or seed-storage proteins are suitable for use in the food industry as modifiers of food properties.

The present inventors have shown that modified proteins according to the present invention can be produced in bacterial fermentation and that large scale production of the proteins for commercial use is possible.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

In order that the present invention may be more clearly understood, preferred forms will be described with reference to the following examples and drawings.

Brief Description of Drawings

Figure 1: Sequence of the pJANGΔCys7Cys236 vector. Restriction sites and enzymes of the MCS of pJKKm and pGEM-T plasmids are written in bold. Cloning sites for gene insertion into ANG are underlined and written in bold. Cleavage sites for sub-cloning the gene into pET 11d expression vector are underlined.

Figure 2: Nucleotide and amino acid sequence of Oleosin Hydrophobic Binding Domain (OHBD). Arrowheads are indicating the direction of primer extension. Smaller letters are indicating extra nucleotides for cloning and amino acids from ANG molecule. Amino acid sequence of the central anti-parallel domain is designated by two arrows.

Figure 3: Nucleotide and amino acid sequence of ANGΔCys7Cys236. (Molecular weight 18.5 kDa) Cysteine residues are marked with bold letters. Extra two amino acids and six nucleotides are written with smaller letters.

Figure 4: Nucleotide and amino acid sequence of the Starch Binding Domain of 1,4-α-D-glucan glucohydrolase from *Aspergillus niger*. (Molecular weight 11.9 kDa). Small letters are indicating extra nucleotides for cloning and amino acids from ANG molecule. Restriction sites are in bold and underlined.

Figure 5: Nucleotide and amino acid sequences of CM16 and CM17. (Molecular weight is 13.4 kDa). Small letters are indicating extra nucleotides for cloning and amino acids from ANG molecule. Restriction sites are in bold and underlined. First amino acid sequence under the nucleotide sequence is representing of CM16 protein, while only differences are shown in CM17 protein.

Figure 6: Nucleotide and amino acid sequence of Puroindoline A. Legend: Molecular weight is 14.3 kDa. Small letters are indicating extra nucleotides for cloning and amino acids from ANG molecule. Restriction sites are in bold and underlined.

Figure 7: Expression of the ANG/OHBD/Cys7Cys236 gene. SDS-PAGE analysis followed by Coomassie blue staining. Lane A contains standard molecular weight markers, lane B shows a control in which the host *E. coli* cell containing the pET-11d vector, and lane C shows the expression of the recombinant ANG/OHBD/Cys7Cys236 protein (migrating slightly further than the 30 kDa marker) from pET-11d containing the ANG/OHBD/Cys7Cys236 gene.

Figure 8: Expression of the ANG/OHBD/Cys7Cys236 gene. SDS-PAGE analysis followed by Western blot using antibodies to C-hordein. Lane A contained standard molecular weight markers, lane B contained extract of an *E. coli* cell expressing a modified C-hordein gene containing a single cysteine in the N-terminal region, and lane C contained extract from *E. coli* containing the ANG/OHBD/Cys7Cys236 gene in the pET-11d expression plasmid.

Figure 9: Expression of the ANG/SBD/Cys7Cys236 gene. SDS-PAGE analysis followed by Western blotting using anti-C-hordein antibodies. Lane B shows the SDS-PAGE profile of standard molecular weight marker proteins. Lanes A and C show the antibody cross reaction of extracts of cells containing either pET-11d vector alone (lane A) or the pET-11d vector containing the ANG/SBD/Cys7Cys236 gene (lane C).

Figure 10: SDS-PAGE analysis of the purified ANG/CM16/Cys7Cys236 gene product by Coomassie blue staining. Lane A contains standard protein molecular weight markers. Lanes B to E show ethanol-soluble extracts of the crude *E. coli* lysates. Lane B contained extract from cells containing the control plasmid, pET-11d. Lanes C, D and E contained ethanol soluble extracts of cells harbouring the pET-11d vector containing the ANG/CM16/Cys7Cys236 gene, prepared from cells 2, 4 and 6 hours respectively after induction of protein synthesis using IPTG.

Figure 11: Expression of the ANG/PIN-A/Cys7Cys236 gene. SDS-PAGE gel stained with Coomassie blue. Lane A contains standard protein molecular weight markers. Lane B contained extract of cells harbouring the control pET-11d plasmid, lanes C and D contained extracts of cells harbouring the pET-11d vector containing the ANG/PIN-A/Cys7Cys236 gene.

Lane C contains extract 2 hours after IPTG induction, lane D contains extract prepared 6 hours following IPTG induction.

Figure 12: Expression of the ANG/PIN-A/Cys7Cys236 gene. SDS-PAGE analysis followed by Western blotting with anti-puroindoline A antibodies. Lanes A and B contained extract of cells harbouring pET-11d containing the ANG/PIN-A/Cys7Cys236 gene, 2 hours and 6 hours after induction respectively. Lane C contains a Western blot of cells harbouring a plasmid containing the puroindoline A gene alone (not inserted into the ANGΔCys7Cys236 gene), and lane D contained extract of cells harbouring the control pET-11d vector.

Figure 13: Expression of the ANG/PIN-A/Cys7Cys236 gene. SDS-PAGE analysis followed by Western blotting with anti-hordein antibodies. Lane A contains molecular weight markers, lane B contained extract of cells harbouring the control pET-11d vector, lane C contained extract of cells harbouring a plasmid which contains the puroindoline A gene alone, and lane D contained extract of cells harbouring pET-11d vector containing the ANG/PIN-A/Cys7Cys236 gene.

Figure 14: Plasmids for the transformation of wheat with constructs containing novel protein genes. The vectors are based on the plasmid pTLZHMWcas which has been modified by insertion of genes into the KpnI and BamH1 site region. The vectors are: pTLZ-ANGCys7Cys13, pTLZ-ANGCys236 and pTLZ_ANG/OHBD/Cys7Cys236.

Modes for Carrying Out the Invention

MATERIALS AND METHODS

Bacterial strains and plasmids

Escherichia coli strain DH5alfa was used as cloning host strain and *E. coli* strain AD494(DE) (Novagen) was used as the expression host in this work.

pGEM-T (Promega) was used as cloning vector of PCR products. pJKKm(-) (Kirschman and Cramer, 1988) served as cloning vehicle for fusion protein genes. and new, assembled genes for expression were sub-cloned into plasmid pET-11d (Novagen).

Cloning of ANGΔCys7Cys236

Restriction endonucleases and DNA modifying enzymes were from New England Biolabs and Promega Corp. Other chemicals and reagents were of analytical reagent grade.

Oligonucleotide primers were synthetised on an Applied Biosystem 394 DNA/RNA Synthetiser using standard phosphoramidite chemistry and were deprotected by heating in ammonium hydroxide solution. Primers were lyophilised and dissolved in TE buffer.

To amplify up a 477 bp long fragment (approximately 2/3 of the whole gene) of C hordein (Accession X60037) from the genomic clone (Lambdahor1-17) of barley by polymerase chain reaction two specially designed primers were used.

oligonucleotide 1:

```
        Met Arg                    Cys
5' GTC ATG AGG CAA CTA AAC CCT TGC AGC CAA GAG TTG CAA TC
3'
     BspH I
```

oligonucleotide 2:

```
             *** Val                   Cys          Gln Gln Pro
5' GGA TCC CTA GAC CAT ACT CCA TAT GCA TGA AGC TTG TTG GGG
    BamH I                  Nde I         Hind III
GAC TGG TTG 3'
```

The reaction was performed in an FTS-4000 Thermal Sequencer (Corbett Research, Australia), with 1 cycle of 3 min at 94°C, 20 s at 55°C, 1 min at 72°C; 36 cycles of 30 s at 94°C, 1 min 30 s at 72°C. The reaction was carried out in 50 µl, containing 45 µl of Supermix (BRL Life Technology) and 1 ng of template DNA and 50 pmol of each oligonucleotides in 5 µl. The DNA was purified following the QIAquick protocol (QIAGEN) and cloned, using the pGEM-T Vector System I (Promega) as was recommended by the manufacturer. White transformant colonies selected for growth in LB medium supplemented with ampicillin (100 mg/l) were screened for insert-bearing plasmid DNA by PCR. Plasmid DNA was purified from positive clones using Jetstar miniprep columns (Genomed) and the insert was sequenced in both direction, using the Prism dye terminator cycle sequencing protocol (Perkin-Elmer). One plasmid isolate, containing the ANG gene, was designated as pGEM-ANG.

This plasmid contained two NdeI restriction sites, one within the gene and one within the pGEM-T multiple cloning site (MCS). To make it unique, the ANG gene was subcloned into pJKKm. Because pJKKm has a HindIII site in MCS, it had to be deleted before subcloning.

One ug of the plasmid was digested with 5 units of HindIII enzyme in 20 µl. Mung Bean Nuclease was used to remove the 5' overhang. Two µl of 10X $ZnSO_4$ solution and 0.02 units of enzyme was added to the reaction mix and incubated at 30^0C. Reaction mix was extracted once with phenol/chloroform after one hour incubation. DNA was recovered with ethanol precipitation and resuspended in 50 µl. An aliquot of 10 µl of the DNA solution was used for ligation. Ligation was carried out in 15 µl solution at 4^0C for overnight using T4 DNA ligase. The ligation mixture was used to transform competent cells of *E. coli* strain DH5alfa by electroporation. Colonies were grown in LB medium supplemented wit kanamycin (50 mg/l). Plasmid was purified from three colonies as mentioned above. DNA samples were tested performing HindIII enzyme digestion. One clone not containing HindIII site was chosen and designated as pJKK-H.

pJKK-H was cut with SphI and BamHI to subclone pGEM-ANG SphI-BamHI fragment into the plasmid. Both digested DNA were purified on QIAquick columns as described earlier, and ligated in 10 µl solution overnight at 14^0C using T4 DNA ligase at a molar ratio of 50:1. Transformed *E. coli* competent cells was spread onto LB medium with kanamycin. Colonies were tested for insert-bearing plasmids DNA by PCR and 3 positive clones were sequenced as above. One clone was designated as pJANGΔCys7Cys236 and used for further cloning work to assemble genes for fusion proteins. Nucleotide and amino acid sequence of this cloning vehicle is shown in Figure 1.

Design and cloning of oleosin hydrophobic binding domain

The sequence of oleosin hydrophobic binding domain (OHBD) was designed to include the consensus sequence of three (maize, rice and barley) oleosin proteins. The sequence is almost identical to the sequence published for barley oleosin isoform-2 (Aalen 1995, Accession Number X82678). Four primers were designed to encode the protein (Figure 2). The gene was constructed by a modification of the technique of overlap extension, where the two long partially overlapping oligonucleotides were further extended and amplified by shorter external primers, encoding restriction enzyme cleavage sites to clone into pJANGΔCys7Cys236. PCR amplification was performed as above, using the same cycle program and Supermix solution. Concentration of the long oligonucleotides were 0.1 nM, while 2 µM of the short primers in the PCR reaction mix. The PCR product was purified on

QIAquick column and cloned into pGEM-T plasmid using Promega kit, as described above. Three positive clones were used for plasmid preparation and sequencing to confirm the nucleotide sequence. One clone was used for further work and called pGEM-OHBD.

Cloning of starch binding domain

The DNA corresponding to the starch binding domain (SBD) of Glucoamylase 1 (1,4-alfa-D-glucan glucohydrolase) of *Aspergillus niger* (Accession number: X00548) was amplified by PCR from purified genomic DNA. Primers were designed to allow us to add tails at both ends of the DNA fragment for cloning into pJANGΔCys7Cys236 vector.

Sequence of the PCR primers are:

oligonucleotide 7:

```
   Gln Ala Cys Thr
5' CAA GCT TGT ACC ACT CCC ACC GCC 3'
    Hind III
```

oligonucleotide 8:

```
       Ile Cys Arg
5' CCA TAT GCA CCG CCA GGT GTC AGT CAC 3'
    Nde I
```

Amplification, cloning into pGEM-T and sequencing was done as described above. One clone bearing the gene fragment was designated as pGEM-SBD.

Cloning of CM16 and CM17 genes for fusion

Both genes were amplified from purified wheat (Triticum aestivum) genomic DNA by PCR.

nucleotide 9:

5' GTC GGC AAT GAA GAT TGC ACC 3'

nucleotide 10:

5' TCC AAC TGC GTT CTC CTC TTG GCC 3'

nucleotide 11:

5' GGA TCC CTA GCT CCA CTG AGA CTC 3'

For CM16 gene (accession number X55455) oligonucleotide 10 and 11, while for CM17 gene (accession number X59791) 9 and 11 pairs were used. Clones were called pGEM-CM16 and pGEM-CM17, respectively. For sub-cloning into pJANGΔCys7Cys236 vector, the genes were PCR amplified again,

using pGEM clones as template, purified on QIAquick column and digested with the appropriate enzymes. Primers were used in this amplification are:
nucleotide 12:

```
            Gln Ala Leu Gly
5' TGC GCT CAA GCT TTA GGC AAT GAA GAT TGC ACC 3'
             Hind III
```

nucleotide 13:

```
                Ile Cys Ser
5'  CAT ACT CCA TAT GCA GCT CCA CTG AGA CTC 3'
              Nde I
```

Cloning puroindoline A gene for fusion

Lambda genomic clone for puroindolin A (PIN-A), kindly provided by Sadequer Rahman, was used as template to amplify the gene (accession number X69913) by PCR. Primers were designed as follows;
oligonucleotide 14:

```
   Gln Ala Tyr
5' CAA GCT TAC GAT GTT GCT GGC GGG 3'
    Hind III
```

oligonucleotide 15:

```
5' CCA TAT GCA CCA GTA ATA GCC AAT AGT GC 3'
    Nde I
```

PCR product was purified, ligated into pGEM-T and sequenced as described above. One clone was used and designated as pGEM-PIN-A.

Fusion of genes or gene fragments with ANG molecule

pJANGΔCys7Cys236 vector was cut with NdeI and HindIII restriction enzymes as all the other pGEM clones and purified on QIAquick columns. Ligation was performed at 14^{0}C overnight in 10 µl of solution containing T4 DNA ligase and insert:vector at about 20:1 molar ratio. One µl of ligation mix was used for transformation of *E. coli* competent cells and spread onto LB plate with kanamycin. Three fusion gene containing colonies were picked up from each transformation for plasmid preparation and sequencing. The clones were called, for example, pJANG/OHBD/Cys7Cys236 in case of the oleosin binding domain containing ANG molecule.

The fused genes were subcloned into pET-11d expression vector between NcoI and BamHI sites. These clones were called pET-ANG/OHBD/Cys7Cys236.

Expression of fusion proteins

For protein expression one of the pET clones were transformed into *E. coli* strain AD494(DE) one day before expression work started.

Small scale expression was carried out in 5 ml 2YT medium, supplemented with Ampicillin (100 mg/l) using one transformant colony. After about 5 hours from inoculation (OD600=0.4), expression was induced to express the protein by addition of 0.4 mM of isopropyl-beta-D-thiogalactopyranoside (IPTG). Both induced and uninduced cultures were further incubated for 4 hours at 37^{0}C. Expression was monitored by SDS-PAGE according to Laemmli (1970).

Large scale expression was performed in a shaking flask. One litre 2YT medium was inoculated with 1 ml of overnight culture and induced to express protein by addition of 0.4 mM IPTG at a cell density of ~0.6 Ab. The culture was incubated with shaking overnight and cells were harvested by centrifugation.

Detection of proteins

PAGE gels were stained overnight with 0.025% Coomassie Blue R-250 in 10% TCA. Excess stain was washed away by water-ethanol-acetic acid (8:1:1) solution.

Immunological detection of the PIN-A containing fusion protein was carried out as it was published earlier (Ciaffi et al., 1998). The antibody was raised against puroindoline crude extract, kindly provided by Sadequr Rahman.

Other chimeric proteins were detected in immunoblott by antibody raised against Hordein, kindly provided by John Skerritt.

Purification of the fusion proteins

Expressed ANG/CM16/Cys7Cys236, ANG/CM17/Cys7Cys236 and ANG/SBD/Cys7Cys236 proteins were purified following the method published elsewhere (Tamas et al., 1998), except for the precipitation step. In this work 2 volumes of 1.5 M NaCl were mixed with the 70% ethanol extract, rather than 4 volumes of acetone.

RESULTS

Design, construction and cloning of pJANGΔCys7Cys236

The gene selected to construct this vector for fusion protein is coding for C hordein. This molecule is a storage protein from barley endosperm and characterised by an absence of cysteine residues. The barley genomic clone

encodes a molecule of 261 residues, including a 20 residue signal peptide. The gene for mature protein (molecular mass 28 kDa) has 723 nucleotides, including a 669 bp long fragment for a central repetitive domain. The oligonucleotides for PCR were designed to reduce the size of the central part and substitute one residue by cysteine in both unique terminal domains. Oligonucleotide 1 binds to the 5' end of the gene and has an additional sequence at the 5' end to incorporate an initiation ATG codon for methionine and a restriction site for BspHI. Oligonucleotide 1 has a T at position 22 to replace an A to change the codon of serine to cysteine at position 7. The 3' end of oligonucleotide 2 is complimentary to a sequence in the repetitive domain between 430 and 448 nucleotides. This is a rather unique sequence within the strongly repetitive region and also codes for an end of a repeat motif. The amplified gene contains only a 411 bp (137 amino acids) long fragment for central repetitive domain. This oligonucleotide also contains the whole sequence of the C-terminal unique region (6 amino acids) of C hordein and a restriction site for BamHI, immediately after the stop codon. To change the threonine residue at the position of 236 of the full size molecule (six residues from the C-terminal end) the oligonucleotide has a C at position 26 and an A at position 27 to substitute a G and a T, respectively. There is another base pair change in this primer at position 22 to replace C with A, which is a "wobble" base of the isoleucine codon. This substitution allowed the creation of restriction site for NdeI enzyme. Oligonucleotide 2 has also got six nucleotides which are not part of the C hordein gene, coding for alanine and serine. These extra nucleotides were added to create one unique restriction site, close to NdeI, within the ANG gene. The two restriction sites are separated with 4 bp to give easy cleavage for both enzymes. The insertion of gene for other molecule or fragment of molecule between HindIII and NdeI enabled the present inventors to create fusion protein with new, designed characteristics.

The gene for ANGΔCys7Cys236 molecule is 474 bp long and codes for a protein with a molecular mass of 18.5 kDa (Figure 3).

PCR amplified DNA was cloned, with two steps, into a plasmid, called pJKKm with a modified MCS. Having deleted Hind III restriction site from the original plasmid, the cloning vehicle pJANGΔCys7Cys236 has unique HindIII and NdeI cleavage sites within ANG gene for gene insertion. Size of the vector is 3873 bp and provides kanamycin resistance to host *E. coli* cells.

The gene is sitting in the vector with 5' ends close to the SP6 RNA Polymerase transcription initiation site.

Subcloning engineered genes into pET-11d did not require agarose gel purification of the fragments, because of the difference in resistance genes within plasmids.

Expression of ANGΔCys7Cys236 protein

To check and compare characteristics of this short molecule to hordeinCys7Cys236, it has been expressed in both small and large quantities. Comparison of the SDS-PAGE patterns of the total cell proteins before and 3 hours after induction with IPTG showed a new band in the induced sample. ANG molecule was readily extracted from lysed cells with 70% (v/v) ethanol and precipitated by the addition of 2 volume of 1.5 M NaCl solution. The resulting preparation, in the presence of reducing agent (0.1 M DTT), had the same mobility as the extra band in *E. coli* lysate. Apparent molecular mass of the protein was about 21 kDa, having a slightly lower mobility in the gel. This characteristic is not unusual for storage proteins. In the absence of DTT, the ethanol extracted sample gave a ladder of bands indicating that the protein is able to form long chains through disulfide bonds.

The ability of ANG molecule to incorporate into gluten matrix of the dough was confirmed by a series of mixing experiments carried out using small scale testers.

Results of SDS-PAGE and mixing experiments showed clearly that ANGΔCys7Cys236 protein had the same or similar properties as the 2 cysteine residue containing C hordein protein.

Design, construction and cloning of synthetic gene for oleosin hydrophobic binding domain (OHBD)

Amino acid sequence was designed according to a comparison of four molecules (one maize, two barleys and one rice), using Genetic Computer Group (GCG) program, called "pileup". OHBD gene fragment, for this work, contain the entire region of the lipophilic stretch of oleosin (Figure 2), very similar to barley gene (accession number: X82677). Codon usage was designed to avoid long stretch of Gs and Cs, which could have led to mis-annealing and sequencing problems. The 5' flanking region of the gene fragment consisted of a tripeptide sequence, of the amphipatic N terminus, while the 3' end another tripeptide of the C terminus of oleosin. The central anti-parallel beta stranded domain had 71 residues. The turn of the anti-

parallel consists of 13 residues and is the most conserved region. The two anti-parallel strands are highly symmetrical in the pairing of residues of similar hydrophobicity on the opposite strands. These characteristics are very similar to those reported for maize oleosin protein (Huang 1996).

The synthetic gene fragment, flanked with two appropriate restriction sites for subcloning, was cloned into pGEM-T vector. One clone, contained the correct sequence of a fragment of 243 bp nucleotides, was designated as pGEM-OHBD. Size of OHBD fragment was 77 amino acids and had a molecular mass of 7 kDa.

Cloning of starch binding domain (SBD)

Glucoamylase 1 from *Aspergillus niger* comprises two domains, one being a catalytic domain (1-470 residues) and the other (509-616 residues) being responsible for binding granular starch (Le Gal-Coeffet et al., 1995).

DNA corresponding to the SBD was amplified by PCR method from *Aspergillus niger* purified genomic DNA, cloned into pGEM-T vector, sequenced and designated as pGEM-SBD. This clone had a 337 bp long fragment (Figure 4) with two restriction sites for insertion into pJANGΔCys7Cys236 vector and an extra alanine before the very first cysteine of SBD. This residue derived from HindIII restriction site. The binding domain had 108 amino acids, two of them are cysteines, with a molecular mass of 11.9 kDa.

Cloning of CM16 and CM17 gene for fusion

Both CM16 and CM17 (CM refers to Chloroform/Methanol soluble) proteins, reported as members of the α–amylase/trypsin inhibitor family and also reported that specific lipids are tightly bound to the fraction, were purified from wheat (Kobrehel and Sauvaire, 1990). These two molecules are very similar on amino acid level, but there are a few differences in the distribution of charged residues (Figure 5).

To clone these two genes, three primers were used to amplify them in two, separate PCR reactions from wheat genomic DNA. One primer (oligonucleotide 11) hybridised to the 3' end of the genes, while to distinguish between the two genes, two specific primers were designed for the 5' ends. pGEM-CM17 clone carried the gene for only the mature protein, however, pGEM-CM16 clone had a few extra base pairs from the signal peptide region. Both clones had a DNA fragment encoding mature

chloroform/methanol soluble proteins with 10 cysteine residues and a molecular mass of 13.4 kDa.

To subclone these genes into ANGΔCys7Cys236 carrying vector, two restriction sites were added to one of each ends by PCR, using the same primer pairs for both pGEM clones, as templates. Primer corresponding to the N-terminus of the mature proteins contained nucleotides for HindIII restriction enzyme. It also had an extra alanine residue, and mutation in the first codon (TTA for leucine), substituting valine (GTC) in CM17 and isoleucine (ATC) in CM16 molecule, because of the sequence requirement for HindIII enzyme. The fragment was cloned into ANG cloning vehicle had 360 nucleotides.

Cloning puroindolin A (PIN-A) gene for fusion

The clone for PIN-A protein, which is capable of binding tightly to both wheat phospholipids and glycolipids (Dubreil et al., 1997), was kindly provided by S. Rahman, derived from a genomic library.

PCR amplified DNA encoded for a 121 amino acid long fragment of puroindoline A, 5 residues shorter in the N terminal region, than the mature protein. This shorter protein had 10 cysteine residues and a molecular mass of 14.3 kDa (Figure 6). For insertion into ANG molecule, the pGEM-PIN-A clone had also carried two restriction sites (HindIII and NdeI) and an extra alanine, because of HindIII site requirement.

Expression of ANG/OHBD/Cys7Cys236 protein

Having inserted OHBD gene fragment into pJANGΔCys7Cys236 cloning vehicle, the entire gene was subcloned into expression vector pET-11d. Protein was expressed in *E. coli*, using AD494(DE) strain. This genetically engineered new gene (717 bp long) coded for a protein with a molecular mass of 25.5 kDa.

Cells were harvested by centrifugation after overnight expression (5 ml culture) and resuspended in 100 μl gel loading buffer (0.125 M TRIS/HCl pH 6.8, 4% SDS, 10% (v/v) glycerol), containing 0.1 M of DTT. Proteins were analysed by SDS-PAGE and stained by Coomassie blue (Figure 7). Lane A contains standard molecular weight markers, lane B shows a control in which the host *E. coli* cell containing the pET-11d vector, and lane C shows the expression of the recombinant ANG/OHBD/Cys7Cys236 protein (migrating slightly further than the 30 kDa marker) from pET-11d containing the ANG/OHBD/Cys7Cys236 gene. The apparent molecular mass of the newly

synthetised protein (~29 kDa) was larger than that calculated on the basis of the nucleotide sequence of DNA. A similar characteristic has been observed previously for different prolamines of cereals. This fusion protein consists of part of a prolamine and a part of an oleosin molecule. C hordein has been considered as a rod shaped molecule, while oleosin hydrophobic binding domain has anti-parallel strands, which would penetrate into the oil body and thus anchor the protein stably. It means that none of the two fragments has globular structure as compared with the molecular weight marker molecules. Discrepancy in apparent molecular mass on SDS-PAGE was probably due to these two structural phenomenon and it was manifested in lower mobility. Figure 8 shows a Western blot of an SDS-PAGE gel (run in the absence of reducing agents such as dithiothreitol) using antibodies to the C-hordein protein. Lane A contained standard molecular weight markers, lane B contained extract of an *E. coli* cell expressing a modified C-hordein gene containing a single cysteine in the N-terminal region, and lane C contained extract from *E. coli* containing the ANG/OHBD/Cys7Cys236 gene in the pET-11d expression plasmid. The strong antibody cross reaction to hordein antibodies at 29 kDa confirms that the 29 kDa protein is the product of the ANG/OHBD/Cys7Cys236 gene.

Expression, purification and analysis of ANG/SBD/Cys7Cys236 protein

A chimeric gene (798 bp in length) for the ANG/SBD/Cys7Cys236 protein was subcloned into pET-11d expression vector and expressed in *E. coli* strain AD494(DE). Cells (4 ml out of 5 ml culture) were harvested 4 hours after induction and resuspended in 200 µl of gel loading buffer with 0.1 M of DTT. One ml of culture was also harvested and the pellet was resuspended in 30 µl water and 70 µl of absolute ethanol was added. Cell suspension was heated at 65^{0}C for 60 min with occasional vigorous shaking. Insoluble cell components were removed by centrifugation at 10,000 g for 15 min. The ethanol-soluble protein was recovered by addition of 2 volume of 1.5 M of NaCl solution. This mixture was stored overnight at -20^{0}C. Insoluble material was collected by centrifugation and the supernatant discarded. Pellet was solubilised in gel loading buffer. Under optimal conditions, 20-25 mg of fusion protein, comprising analogue glutenin and starch binding domain, was obtained in large scale expression, from 1 l of bacterial cell culture.

Samples of crude were analysed by Western blotting following SDS-PAGE, using a monoclonal anti-hordein antibody. The result is shown in Figure 9. Lane B shows the SDS-PAGE profile of standard molecular weight marker proteins. Lanes A and C show the antibody cross reaction of extracts of cells containing either pET-11d vector alone (Lane A) or the pET-11d vector containing the ANG/SBD/Cys7Cys236 gene (lane C). The antibody cross-reacted with a range of proteins in each sample, however, a novel immunoreactive band is clearly seen in extracts from cells containing the ANG/SBD/Cys7Cys236 gene (lane C). The apparent molecular mass of the newly synthetised chimeric protein (~36 kDa) was larger than was calculated according to the amino acid sequence (30.4 kDa). Lower mobility can be explained by the special shape of the molecule (as defined above).

Expression, purification and analysis of ANG/CM16/Cys7Cys236 protein

Gene for this fusion protein (834 bp long) was subcloned, from kanamycin resistant, pJKKm originated plasmid, into the ampicillin resistant pET-11d expression vector, as above. Bacteria was grown at 37^{0}C until OD600 reached 0.6 units, then expression was induced by adding IPTG. Cells were harvested two, four or six hours after induction and samples were prepared as in case of SBD chimeric protein.

Result of the SDS-PAGE analysis is shown in Figure 10 (the gel is stained with Coomassie protein stain). Lane A contains standard protein molecular weight markers. Lanes B to E show ethanol-soluble extracts of the crude *E. coli* lysates. Lane B contained extract from cells containing the control plasmid, pET-11d. Lanes C, D and E contained ethanol soluble extracts of cells harbouring the pET-11d vector containing the ANG/CM16/Cys7Cys236 gene, prepared from cells 2, 4 and 6 hours respectively after induction of protein synthesis using IPTG. The ethanol-soluble fraction gave only one band on Coomassie stained gel representing a protein with an apparent molecular mass of 35 kDa. This was also higher than the calculated (31.9 kDa), and this discrepancy is also explained by the unusual shape of the protein.

Yield of protein expression in large scale, in shaking flask was 30 mg protein from 1 litre of medium.

Expression and analysis of ANG/PIN-A/Cys7Cys236 protein

A DNA fragment comprised of ANGΔCys7Cys236, and Puroindoline-A genes with a length of 840 nucleotides from kanamycin resistant plasmid (see

above) was subcloned into pET-11d expression vector. One transformant colony was picked up and transferred into 5 ml of 2YT medium. Protein expression was induced by IPTG and after o2 and 6 hours cells were harvested by centrifugation and resuspended in 200 µl of gel loading buffer. SDS-PAGE analysis of protein content of bacteria, before and after induction, is shown on Figure 11. Lane A contains standard protein molecular weight markers. Lane B contained extract of cells harbouring the control pET-11d plasmid, lanes C and D contained extracts of cells harbouring the pET-11d vector containing the ANG/PIN-A/Cys7Cys236 gene. Lane C contains extract 2 hours after IPTG induction, lane D contains extract prepared 6 hours following IPTG induction. An extra protein band appeared on gel with an apparent molecular mass of 35 kDa, after induction (lane D). This expressed protein positively reacted with puroindoline A antibody (Figure 12). Lanes A and B contained extract of cells harbouring pET-11d containing the ANG/PIN-A/Cys7Cys236 gene, 2 hours and 6 hours after induction respectively. Lane C contains a Western blot of cells harbouring a plasmid containing the puroindoline A gene alone (not inserted into the ANGΔCys7Cys236 gene), and lane D contained extract of cells harbouring the control pET-11d vector. All lanes were reacted with anti- puroindoline A antibody. Figure 13 shows a Western blot of an SDS-PAGE gel reacted with anti-hordein antibodies. Lane A contains molecular weight markers, lane B contained extract of cells harbouring the control pET-11d vector, lane C contained extract of cells harbouring a plasmid which contains the puroindoline A gene alone, and lane D contained extract of cells harbouring pET-11d vector containing the ANG/PIN-A/Cys7Cys236 gene. The expressed protein is clearly seen in lane D.

Analogue glutenin-puroindoline A fusion protein had a slightly lower mobility on SDS-PAGE as it was calculated (32.8 kDa) on the basis of amino acid sequence. This small difference was probably due to the rod shape of the ANG part of the molecule. Expressed puroindoline A protein, however, had the same mobility as it was calculated.

Expression of modified proteins following wheat transformation.

Microprojectile bombardment is currently the most widely applied technique used to transfer genes into wheat. The transfer of genes into wheat and their expression is carried out using specific DNA constructs containing a selectable marker gene and the gene-of-interest respectively. The selectable

marker gene used was contained on the plasmid pEmuKON (Chamberlain et al., 1994). Constructs for the gene-of-interest constructs are shown in Figure 14. Wheat transformation was carried out using paromomycin selection according to the procedures outlined in Witrzens et al., 1998).

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Dated this tenth day of July 1998

COMMONWEALTH SCIENTIFIC
AND INDUSTRIAL RESEARCH
ORGANISATION,
GOODMAN FIELDER LIMITED,
GROUPE LIMAGRAIN PACIFIC
PTY LTD
Patent Attorneys for the Applicant:

F B RICE & CO

REFERENCES

Aalen, R.B. 1995. The transcript encoding two oleosin isoforms are both present in the aleurone and in the embryo of barley (Hordeum vulgare) seeds. Plant Mol. Biol. 28:583-588

Bekes, F., Gras, P.W. and Gupta, R.B. 1994. Mixing properties as a measure of reversible reduction and oxidation of doughs. Cereal Chemistry 71:44-50

Bushuk, W. 1998 . Interactions in wheat doughs. *In* Interactions: The keys to cereal quality. R.J. Hamer, R.C. Hoseney Eds. American Association of Cereal Chemists, Inc. St Paul, Minnesota. pp1-16.

Chamberlain, D.A., Brettell, R.I.S., Last, D.I., Witrzens, B., McElroy, D., Dolferus, R., Dennis, E.S. 1994. The use of the EMu promoter with antibiotic and herbicide resistance genes for the selection of transgenic wheat callus and rice plants. Australian Journal of Plant Physiology. 21: 95-112.

Ciaffi, M., Lee, Y.K., Tamas, L., Gupta, R., Skerritt, J. and Appels, R. (in press) Molecular characterization of LMW glutenin subunit genes in the D genome donor of hexaploid wheat (Triticum tauschii). Theoretical and Applied Genetics

Dubreil, L., Compoint, J-P. and Marion, D. 1997. Interaction of Puroindolines with wheat flour polar lipids determines their foaming properties. J. Agric. Food Chem. 45:108-116

Gan, Z., Ellis, P.R. and Schofield, J.D. 1995. Gas cell Stabilisation and gas retention in wheat bread dough. Journal of Cereal Science 21:215-230

Huang, A.H.C. 1996. Oleosins and oil bodies in seeds and other organs. Plant Physiology 110:1055-1061

Kasarda, D.D. 1989. Glutenin structure in relation to wheat quality. Pages 277-302 in : Wheat is unique. Y. Pomeranz, Ed. Amer. assoc. Cereal Chem., St Paul, MN.

Kobrehel, K. and Sauvaire, Y. 1990. Particular lipid composition in isolated proteins of durum wheat. J. Agric. Food Chem. 38:1164-1171

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685

Le Gal-Coeffet, M-F., Jacks, A.J., Sorimachi,K., Williamson, M.P., Williamson, G. and Archer, D.B. 1995. Expression in Aspergilus Niger of the starch-binding domain of glucoamylase. Eur. J. Biochem. 233:561-567

MacRitchie, F. 1992. Physicochemical properties of wheat proteins in relation to functionality. Adv. Food Nutr. Research 36:1-87

Morrison, W.R. 1989. Recent progress on the chemistry and functionality of flour lipids. Pages 131-149 in: Wheat end-use properties: Wheat and flour characterisation for specific end-uses. H. Salovaara, Ed. University of Helsinki, Lahti.

Tamas, L., Bekes, F., Greenfield, J., Tatham, A.S., Gras, P.W., Shewry, P.R. and Appels, R. 1998. Heterologous expression and dough mixing studies of wild-type and mutant C hordeins. Journal of Cer. Science 27:15-22

Weeks, J.T. ,Anderson, O.D. and Blechl, A.E. 1993. Rapid production of multiple independent lines of fertile transgenic wheat (Triticum aestivum). Plant Physiology 102:1077-1084

Witrzens, B., Brettell, R.I.S., Murray, F.R., McElroy, D., Li, Z., Dennis, E.S. 1998. Comparison of three selectable marker genes for transformation of wheat by microprojectile bombardment.

Sequence of pJANGΔCys7Cys236 vector

```
                                                                M   R   Q   L   N   P   C   S
5'... ATAGAATAC A GCATGC TCC CGGCCG CCATGG CCGCGG GATTGTC ATG AGG CAA CTA AAC CCT TGC AGC....
  SP6 Promoter<|     SphI       EagI   NcoI   SacII        BspHI

                      V   P   Q   Q   A   S   C   I   W   S   M   V  ***
ANG SEQUENCE....GTC CCC CAA CAA GCT TCA TGC ATA TGG AGT ATG GTC TAG GGATCC
                            HindIII       NdeI                  BamHI

GGGTACC GAGCTC GAATTC GCCCTATA... 3'
 KpnI    SacI    EcoRI |> T7 Promoter
```

Figure 1

Nucleotide and amino acid sequence of Oleosin Hydrophobic Binding Domain

(Synthetic gene)

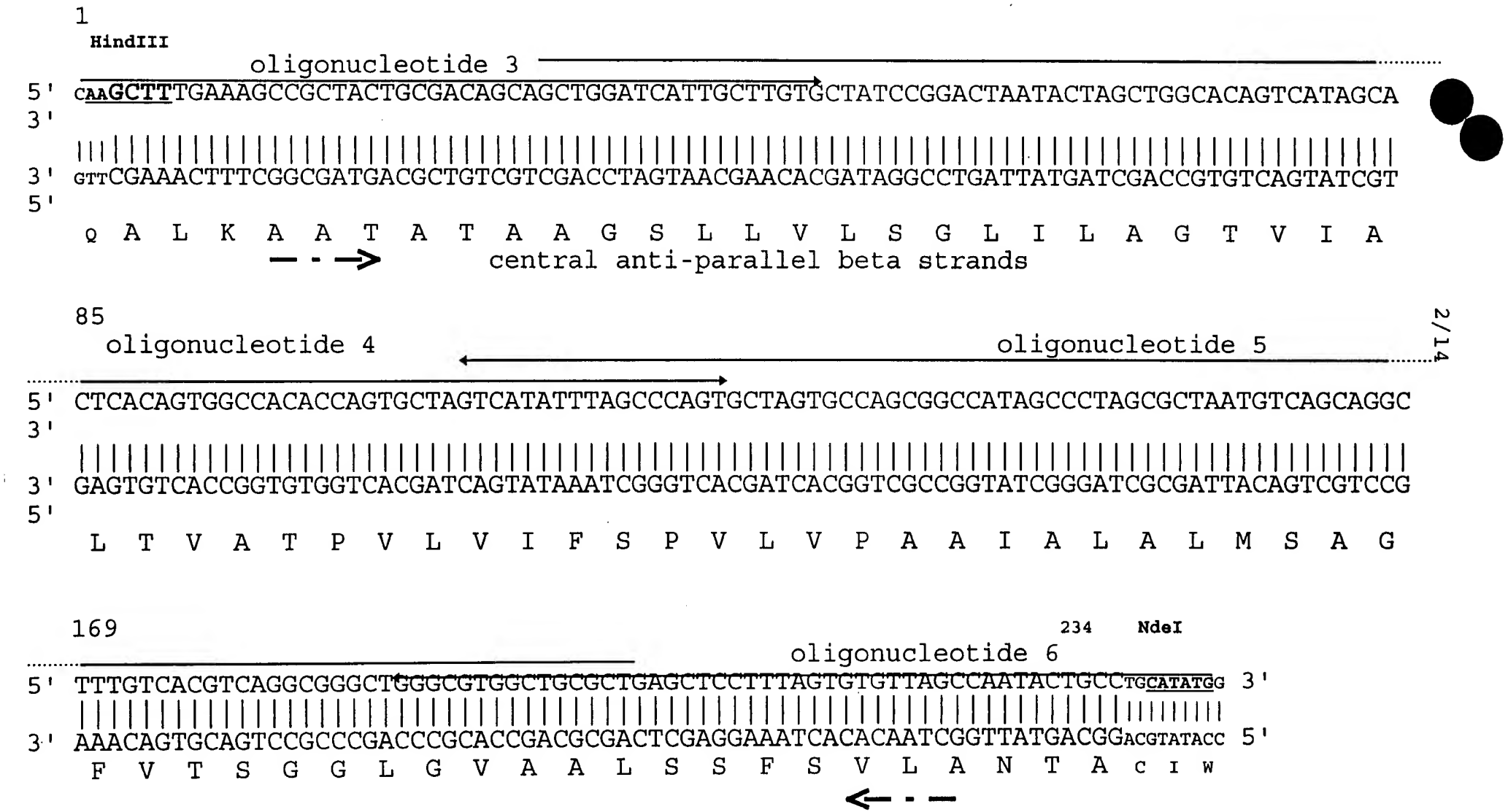


Figure 2

Nucleotide and amino acid sequence of ANGΔCys7Cys236

1
ATGAGGCAACTAAACCCT**TGC**AGCCAAGAGTTGCAATCACCACAACAATCATATCTGCC
G
M R Q L N P **Cys** S Q E L Q S P Q Q S Y L Q

61
CAGCCATATCCACAAAACCCATATCTACCGCAAAAACCATTTCCAGTGCAGCAACCGTT
T
Q P Y P Q N P Y L P Q K P F P V Q Q P F

121
CACACACCCCAACAATATTTCCCCTATCTACCAGAGGAATTGTTTCCCCAATATCAAAT
A
H T P Q Q Y F P Y L P E E L F P Q Y Q I

181
CCAACCCCCCTACAACCACAACAACCATTCCCCCAACAACCACAACAACCTCTTCCTCG
G
P T P L Q P Q Q P F P Q Q P Q Q P L P R

241
CCCCAACAACCATTCCCCTGGCAACCACAACAACCATTTCCCCAGCCCCAAGAACCAAT
T
P Q Q P F P W Q P Q Q P F P Q P Q E P I

301
CCCCAGCAACCACAACAACCATTCCCACAGCAACCACAACAACCATTCCCACAGCAACC
A
P Q Q P Q Q P F P Q Q P Q Q P F P Q Q P

361
CAACAAATAATTTTCCAGCAACCCCAACAATCATACCCTGTGCAACCTCAACAGCCATT
T
Q Q I I F Q Q P Q Q S Y P V Q P Q Q P F

421
477
CCTCAACAACCTCAACCAGTCCCCCAACAA GCT TCA **TGC**ATATGGAGTATGGTCTAG
P Q Q P Q P V P Q Q A S **Cys** I W S M V ***

Figure 3

Nucleotide and amino acid sequence of the Starch Binding Domain of 1,4-α-D-glucan glucohydrolase from *Aspergilus niger*

HindIII 1 54

AAGCTTCTACCACTCCCACCGCCGTGGCTGTGACTTTCGATCTGACAGCTACCACCACCTAC

A S T T P T A V A V T F D L T A T T T Y

114

GGCGAGAACATCTACCTGGTCGGATCGATCTCTCAGCTGGGTGACTGGGAAACCAGCGAC

G E N I Y L V G S I S Q L G D W E T S D

174

GGCATAGCTCTGAGTGCTGACAAGTACACTTCCAGCGACCCGCTCTGGTATGTCACTGTG

G I A L S A D K Y T S S D P L W Y V T V

234

ACTCTGCCGGCTGGTGAGTCGTTTGAGTACAAGTTTATCCGCATTGAGAGCGATGACTCC

T L P A G E S F E Y K F I R I E S D D S

294

GTGGAGTGGGAGAGTGATCCCAACCGAGAATACACCGTTCCTCAGGCGTGCGGAACGTCG

V E W E S D P N R E Y T V P Q A C G T S

321 NdeI

ACCGCGACGGTGACTGACACCTGGCGGTGCATATGG

T A T V T D T W R C I W

Figure 4

Nucleotide and amino acid sequences of CM16 and CM17

HindIII

```
                                                                      57
AAGCTTCGGCAATGAAGATTGCACCCCATGGATGAGTACTCTGATCACTCCACTCCCAAGC
  A I  G  N  E  D  C  T  P  W  M  S  T  L  I  T  P  L  P  S
CM17. .  .  .  .  .  .  .  .  T  .  .  .  .  .  .  .  .  .

                                                                     117
TGCCGTGACTATGTGGAACAACAAGCATGTCGCATCGAAACGCCCGGGTCGCCGTACCTC
 C  R  D  Y  V  E  Q  Q  A  C  R  I  E  T  P  G  S  P  Y  L
 .  .  N  .  .  .  E  .  .  .  .  .  .  M  .  .  P  .  .  .

                                                                     177
GCCAAGCAGCAGTGCTGTGGGGAGCTTGCAAACATTCCGCAGCAGTGCCGATGCCAGGCG
 A  K  Q  Q  C  C  G  E  L  A  N  I  P  Q  Q  C  R  C  Q  A
 .  .
 .  E  .  .  E  Q  .  .  .  .  .  .  .  .  .  .  .

                                                                     237
CTGCGCTACTTCATGGGGCCGAAGTCTCGTCCGGATCAGAGCGGCCTCATGGAACTCCCC
 L  R  Y  F  M  G  P  K  S  R  P  D  Q  S  G  L  M  E  L  P
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

                                                                     297
GGATGCCCTAGGGAGGTGCAGATGGACTTCGTGAGGATACTCGTCACGCCGGGGTACTGC
 G  C  P  R  E  V  Q  M  D  F  V  R  I  L  V  T  P  G  Y  C
 .  .  .  .  .  .  .  .  N  .  .  P  .  .  .  .  .  .  .  .

                                                                    354
AACTTGACGACCGTTCACAACACTCCGTACTGCCTCGCTATGGAGGAGTCTCAGTGG
 N  L  T  T  V  H  N  T  P  Y  C  L  A  M  E  E  S  Q  W
 .  .  .  .  .  .  .  .  .  .  .  .  G  .  .  .  .  .  .

357 NdeI
AGCTGCATATGG
 S  C  I  W
 .
```

Figure 5

Nucleotide and amino acid sequence of Puroindoline A

HindIII

AAGCTTACGATGTTGCTGGCGGGGGGTGGTGCTCAACAATGCCCTGTAGAGACAAAGCTAAAT
A Y D V A G G G G A Q Q C P V E T K L N

TCATGCAGGAATTACCTGCTAGATCGATGCTCAACGATGAAGGATTTCCCGGTCACCTGG
S C R N Y L L D R C S T M K D F P V T W

CGTTGGTGGAAATGGTGGAAGGGAGGTTGTCAAGAGCTCCTTGGGGAGTGTTGCAGTCC
R W W K W W K G G C Q E L L G E C C S R

CTCGGCCAAATGCCACCGCAATGCCGCTGCAACATCATCCAGGGGTCAATCCAAGGCGAT
L G Q M P P Q C R C N I I Q G S I Q G D

CTCGGTGGCATCTTCGGATTTCAGCGTGATCGGGCAAGCAAAGTGATACAAGAAGCCAAG
L G G I F G F Q R D R A S K V I Q E A K

AACCTGCCGCCCAGGTGCAACCAGGGCCCTCCCTGCAACATCCCCGGCACTATTGGCTAT
N L P P R C N Q G P P C N I P G T I G Y

363 NdeI
TACTGGTGCATATGG
Y W C I W

Figure 6

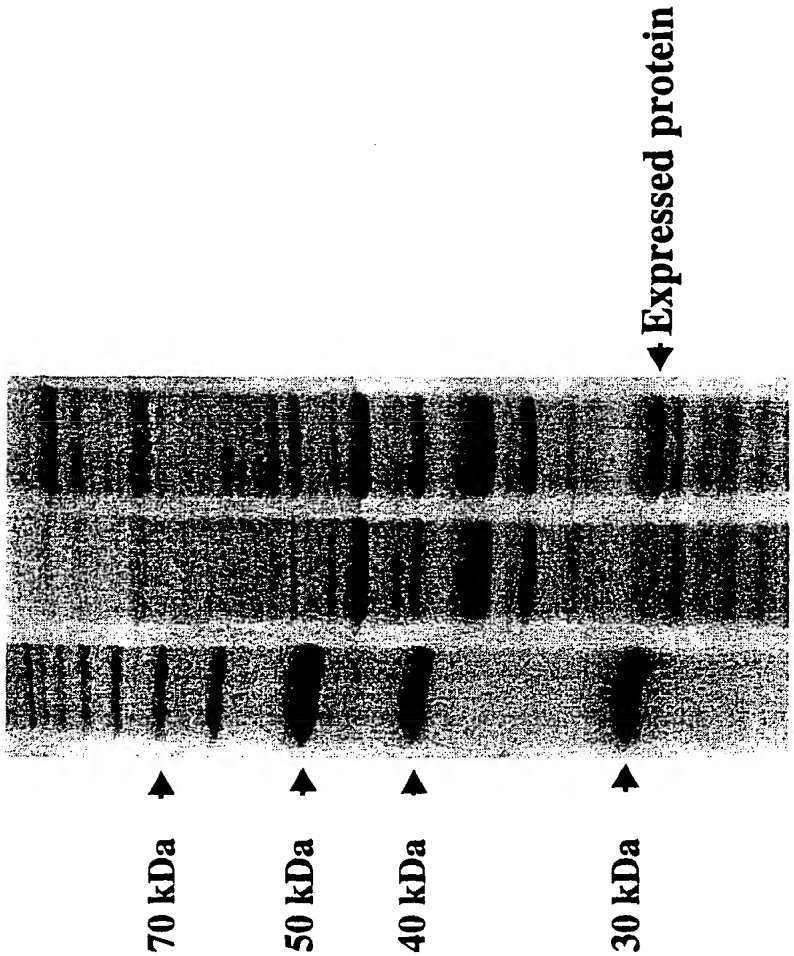


Figure 7

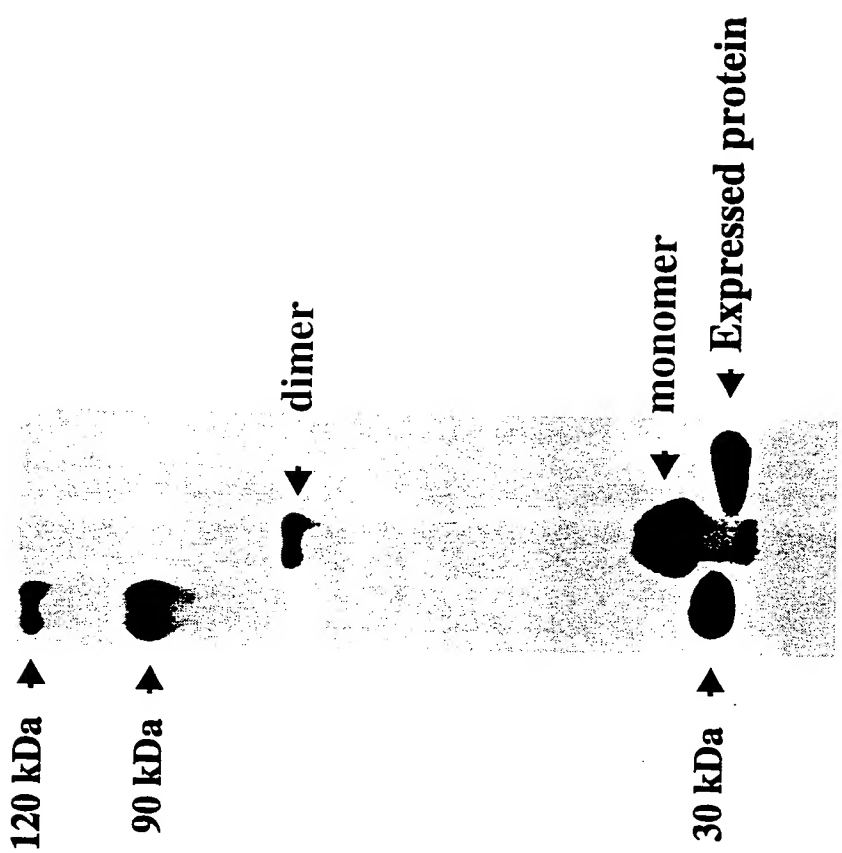


Figure 8

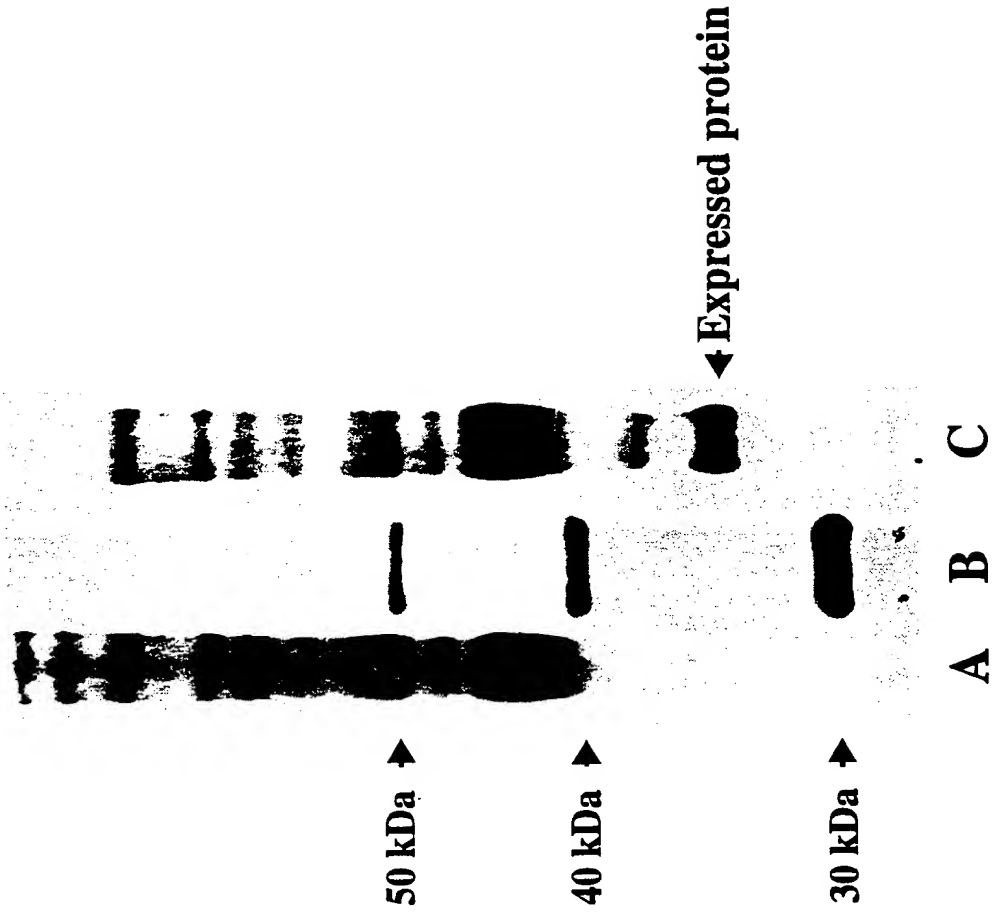


Figure 9

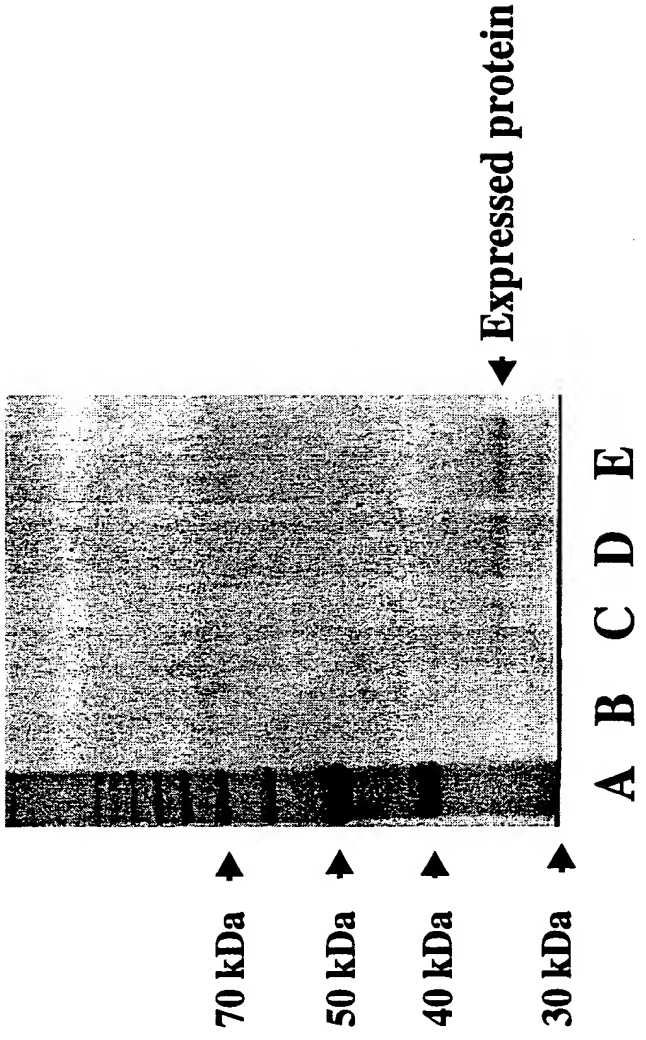


Figure 10

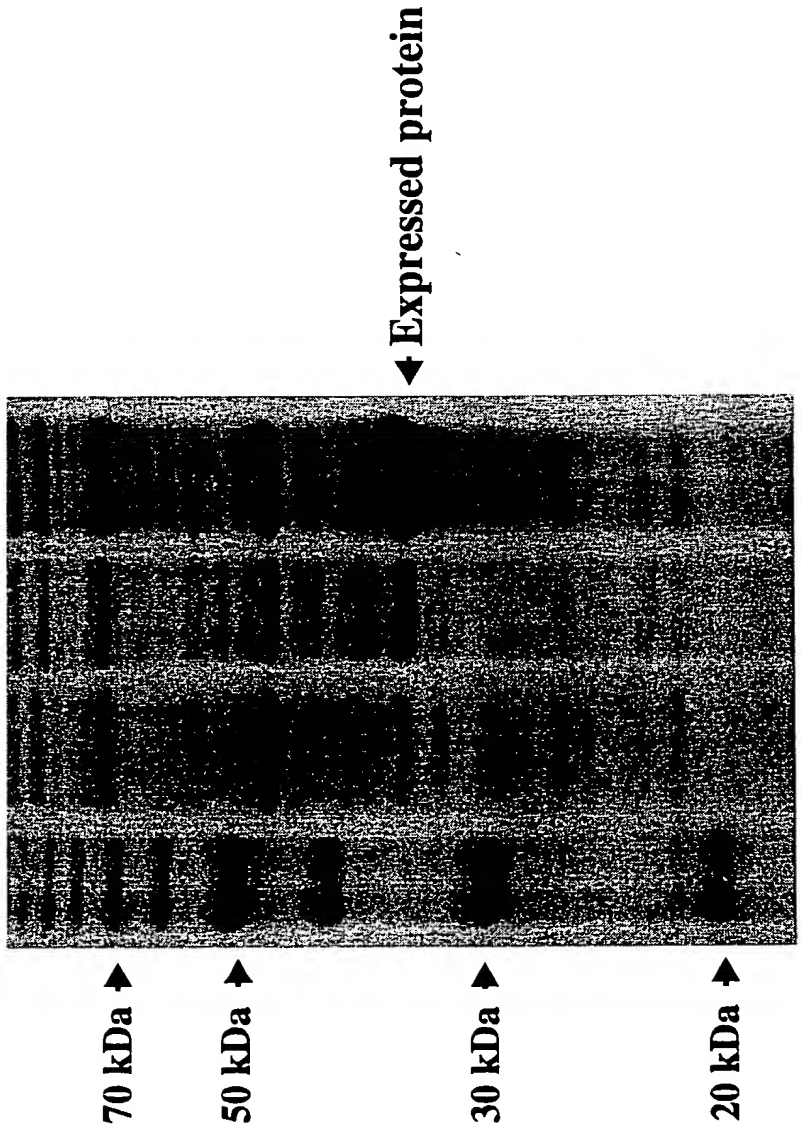


Figure 11

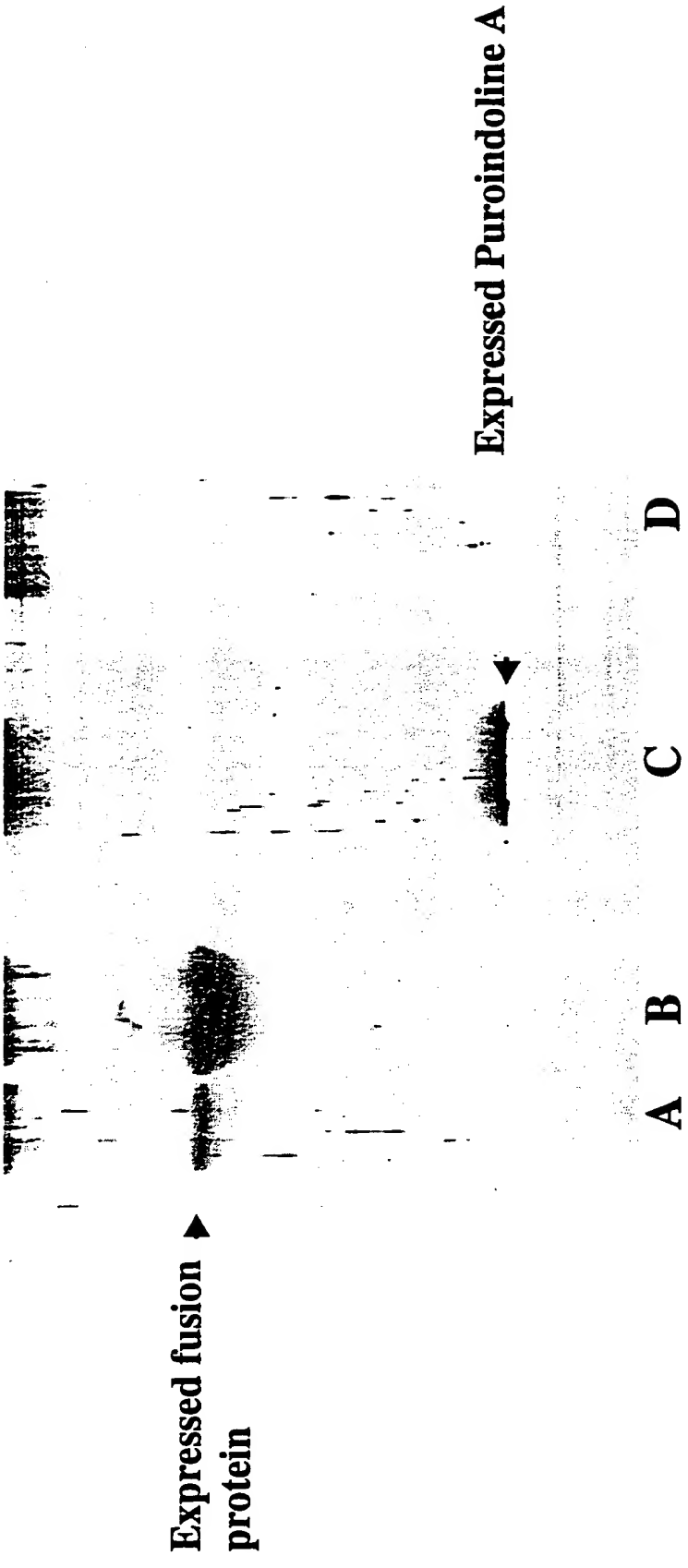


Figure 12

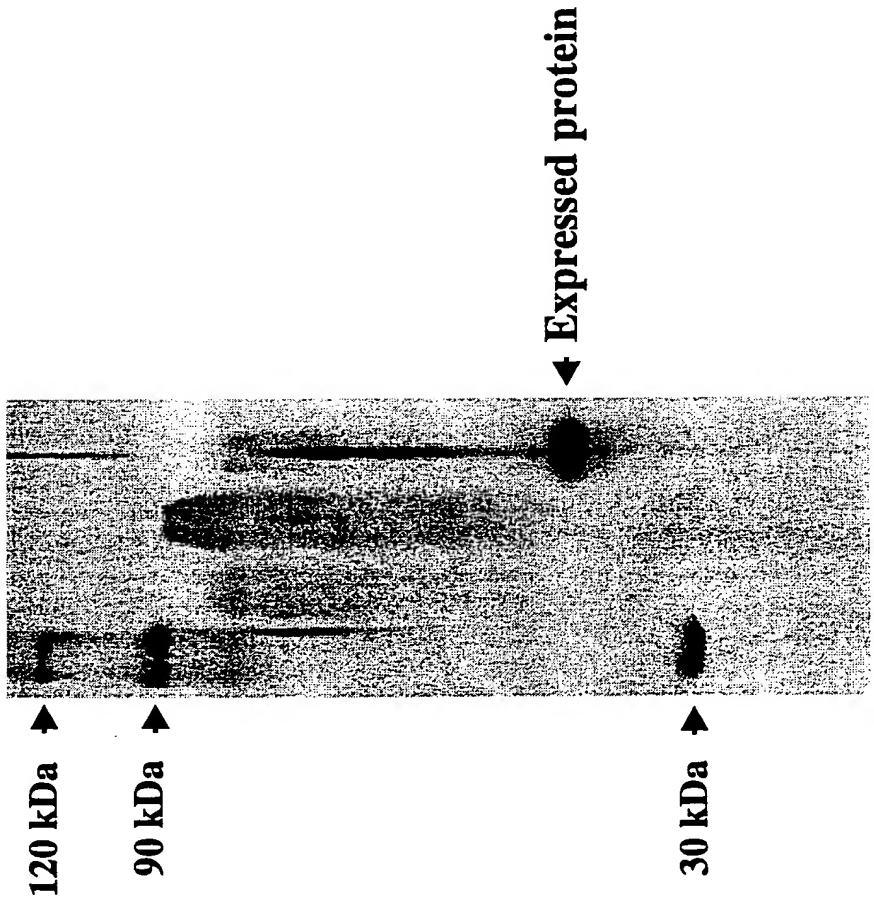
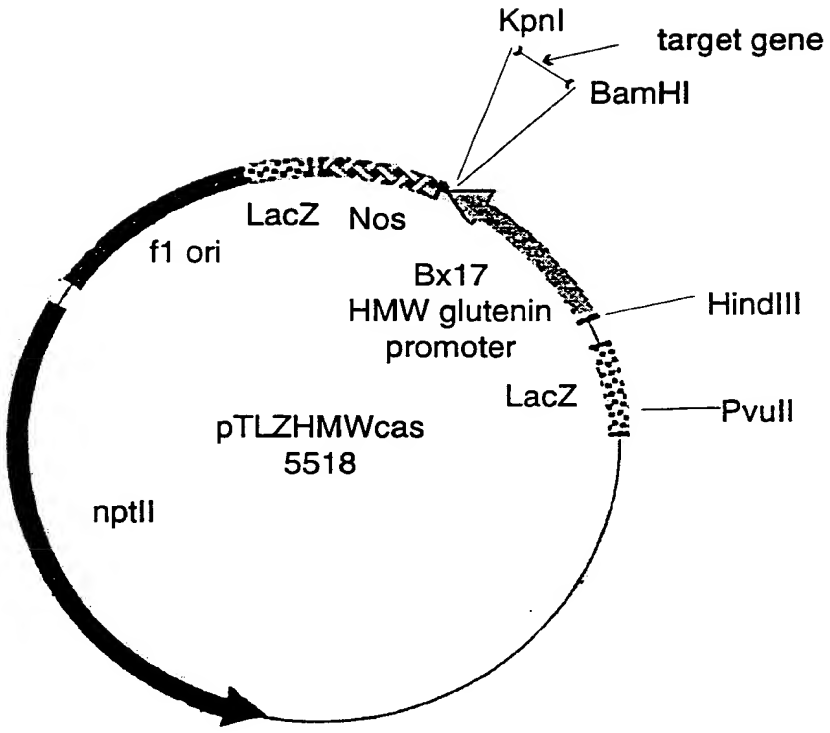


Figure 13



Constructs

(1) pTLZ_ANGCys7Cys13. Contains the C-Hordein 20 amino acid signal peptide + ANGCys7Cys13
(2) pTLZ_ANGCys236. Contains the C-Hordein 20 amino acid signal peptide + ANGCys236
(3) pTLZ_ANG/OHBD/Cys7Cys236. Ontains the C-Hordein 20 amino acid signal peptide + ANG/OHBD/Cys7Cys236

Figure 14